KATHOLIEKE UNIVERSITEIT
LEUVEN

Faculty of Business and Economics

# Goodness-of-fit tests for the frailty distribution in proportional hazards models with shared frailty

Candida Geerdens, Gerda Claeskens and Paul Janssen

DEPARTMENT OF DECISION SCIENCES AND INFORMATION MANAGEMENT (KBI)

KBI 1131

# Goodness-of-fit tests for the frailty distribution in proportional hazards models with shared frailty

**Candida Geerdens[1,*], Gerda Claeskens[2,**], and Paul Janssen[1,***]**

[1]Center for Statistics, Hasselt University, Agoralaan Building D, B-3590 Diepenbeek, Belgium,

[2]ORSTAT and Leuven Statistics Research Center, K.U.Leuven, Naamsestraat 69, B-3000 Leuven, Belgium

**email:* Candida.Geerdens@UHasselt.be

***email:* Gerda.Claeskens@econ.kuleuven.be

****email:* Paul.Janssen@UHasselt.be

SUMMARY: Frailty models account for the clustering present in grouped event time data. A proportional hazards model with shared frailties expresses the hazard for each subject. Often a one-parameter gamma distribution is assumed for the frailties. The choice of a particular frailty distribution is, most of the time, based on the availability of software, rather than on the way it fits the data. In this paper we construct formal goodness-of-fit tests to test for gamma frailties. We construct a new class of frailty models that extend the gamma frailty model by using certain polynomial expansions that are orthogonal with respect to the gamma density. For this extended family we obtain an explicit expression for the marginal likelihood of the data. The order selection test is based on finding the best fitting model in such a series of expanded models. A bootstrap is used to obtain p-values for the tests. Simulations and data examples illustrate the test's performance.

Keywords: Gamma distribution; Goodness-of-fit test; Frailty model; Order selection test; Orthogonal polynomial.

## 1. Introduction

In many experimental settings event data arise in groups (clustered data). As an illustration consider the following two examples from veterinary medicine.

In mastitis studies, times to infection of udder quarters are observed. Since each udder quarter is separated from the three other quarters, one quarter might be infected with the other quarters infection free. The grouping is at the cow level; the cluster size is four (Goethals et al., 2009).

In artificial insemination programmes for dairy cows, times from parturition to first insemination are observed as this is one of the main factors determining the length of the calving interval that lies between 12 and 13 months (Duchateau et al., 2005). For such data the clustering is at the dairy farm level; the cluster size is typically large and varies per cluster.

The mastitis data and the artificial insemination data will be used to illustrate the goodness-of-fit methodology proposed in this paper.

Frailty models account for the clustering present in grouped event time data. For a subject $j$ ($j = 1, \ldots, n_s$) in cluster $s$ ($s = 1, \ldots, S$) denote $Y_{sj} = \min\{T_{sj}, C_{sj}\}$, where $T_{sj}$ is the event time for this subject, and $C_{sj}$ is the censoring time. The indicator $\delta_{sj} = I(T_{sj} \leqslant C_{sj})$ is one for a subject where the event has taken place, while $\delta_{sj} = 0$ for a censored observation. Event times and censoring times are assumed to be independent.

A proportional hazards model with shared frailties expresses the hazard for subject $j$ in cluster $s$ at time $t$ by

$$h_{sj}(t) = h_0(t) \exp(x_{sj}^T \beta + W_s) = h_0(t) U_s \exp(x_{sj}^T \beta), \tag{1}$$

where $h_0(t)$ is the baseline hazard function and $U_s = \exp(W_s)$ is the frailty for cluster $s$ that is common to all members of that cluster. The frailties $U_1, \ldots, U_S$ are assumed to be independent and identically distributed copies of a generic frailty $U$. A commonly used distribution for $U$ is the one-parameter gamma distribution, with mean equal to one, of

which the density function takes the form $f_U(u) = u^{1/\theta-1}e^{-u/\theta}\theta^{-1/\theta}/\Gamma(1/\theta)$ with $\Gamma(\cdot)$ the gamma function. Other often used distributions for $U_s$ are inverse Gaussian and positive stable distributions. For detailed information about frailty models, we refer to the book by Duchateau and Janssen (2008). The choice of a particular frailty distribution is, most of the time, based on the availability of software, rather than on the way it fits the data.

In this paper we address testing the null hypothesis of a gamma frailty distribution. We construct tests that are nonparametric in spirit and that are in the style of the order selection tests for the normality of random effects in linear mixed models (Claeskens and Hart, 2009), using ideas of flexible modeling of random effects by series expansions inspired by Zhang and Davidian (2001) (see also Gallant and Nychka, 1987). The estimators and test statistics in those papers were constructed for testing for normality of the random effect distribution, it is therefore natural to consider Hermite expansions. Indeed Hermite polynomials possess orthogonality properties with respect to the normal distribution. To test for the gamma frailty distribution, other choices of basis expansions should be considered. In particular, we construct a new class of frailty models that extend the gamma frailty model by using certain polynomial expansions that are orthogonal with respect to the gamma density. For this extended family we obtain an *explicit* expression for the marginal likelihood of the data.

Nonparametric order selection tests (Eubank and Hart, 1992; Hart, 1997) are naturally phrased for estimators defined via series expansions. A data-driven model selection method such as an adapted version of Akaike's information criterion AIC (Akaike, 1973) is used to find the truncation point of the series. The test rejects the null hypothesis when at least one term is selected. Aerts et al. (1999, 2000) obtained the asymptotic distribution of the order selection tests for testing parametric hypotheses in likelihood regression models with independent observations. Our work extends some of these ideas to gamma frailty models

and is more in line with the order selection tests in linear mixed models (Claeskens and Hart, 2009).

## 2. A class of extended gamma frailty densities

For identifiability reasons (see below), we standardize the frailties such that these have expectation equal to one. Denote $U_m = \tilde{U}_m / E(\tilde{U}_m)$ where $\tilde{U}_m$ is the unstandardized frailty for the model indexed by $m$. Instead of assuming that the frailties have a one-parameter gamma density we assume that the frailties $\tilde{U}_m$ have a density of form

$$f_{\tilde{U}_m}(u) = \frac{f_U(u)}{c(d)} \{\sum_{j=0}^{m} d_j v_j(u)\}^2 \tag{2}$$

where $f_U$ is the one-parameter gamma density, $v_j$ are polynomials orthonormal to $f_U$, defined in Lemma 1, and $c(d) = d^T d = \sum_{j=0}^{m} d_j^2$ is a normalization constant, with $d_0 = 1$ and $d = (d_0, \ldots, d_m)$. Note that $m = 0$ corresponds to the regular one-parameter gamma density and that $U_0 = U$ (the null model). By fitting models with different values for the truncation point $m$ and using a model selection method to determine the appropriate value of $m$ one can test whether or not the null model describes the data accurately.

Similar to the generalized Laguerre polynomials that are orthogonal with respect to the weight function $u^a \exp(-u)$, it follows via an application of the Gram-Schmidt orthogonalization procedure (the proof is straightforward and provided in Web Appendix A) that the following polynomial functions $v_n$ are orthonormal with respect to the density $f_U$.

LEMMA 1: *For $n = 0, 1, 2, \ldots$, the polynomial functions $p_n : \mathbb{R}^+ \to \mathbb{R} : u \to p_n(u)$ with*

$$p_n(u) = (-\theta)^n u^{-\left(\frac{1}{\theta}-1\right)} e^{\frac{u}{\theta}} \frac{d^n}{du^n} \left[ e^{-\frac{u}{\theta}} u^{\frac{1}{\theta}+n-1} \right] = \sum_{i=0}^{n} (-\theta)^{n-i} \binom{n}{i} \frac{\Gamma\left(\frac{1}{\theta}+n\right)}{\Gamma\left(\frac{1}{\theta}+i\right)} u^i$$

*are orthogonal with respect to the inner product $\langle g_1, g_2 \rangle = \int_0^\infty g_1(u) g_2(u) f_U(u) du$, where $f_U$ is the one-parameter gamma density function. The set of functions $\{v_n; n = 0, 1, 2, \ldots\}$ is orthogonal, where $v_n(u) = p_n(u)/\|p_n\|^{1/2}$ with $\|p_n\| = n! \theta^{2n} \Gamma\left(\frac{1}{\theta}+n\right) / \Gamma\left(\frac{1}{\theta}\right)$.*

Concretely, $v_0(u) = 1$, $v_1(u) = (u-1)\theta^{-1/2}$, $v_2(u) = \{u^2 - u(2+2\theta) + \theta + 1\}/\{\theta(2+2\theta)^{1/2}\}$, $v_3(u) = \{u^3 - 3(1+2\theta)u^2 + 3(1+3\theta+2\theta^2)u - (1+3\theta+2\theta^2)\}/\{6\theta^3(1+3\theta+2\theta^2)\}^{1/2}$. A graphical representation is given in Figure 1.

[Figure 1 about here.]

We define the coefficients $a_{ij} = (-\theta)^{j-i}\binom{j}{i}\Gamma\left(\frac{1}{\theta}+j\right)/\Gamma\left(\frac{1}{\theta}+i\right)$, $a^*_{ij} = a_{ij}/\|p_j\|^{1/2}$, $b^*_{ij} = a^{*2}_{ij} + 2\sum_{k+l=2i,k<l\leqslant j} a^*_{kj}a^*_{lj}$, and $c^*_{ij} = 2\sum_{k+l=2i+1,k<l\leqslant j} a^*_{kj}a^*_{lj}$.

LEMMA 2: *For $\tilde{U}_m$ a random variable with density $f_{\tilde{U}_m}$ as in* (2),

$$E_m = E(\tilde{U}_m) = \Big[\sum_{j=0}^{m} d_j^2\Big\{\sum_{i=0}^{j} b^*_{ij}\theta^{2i+1}\Gamma(2i+1/\theta+1) + \sum_{i=0}^{j-1} c^*_{ij}\theta^{2i+2}\Gamma(2i+1/\theta+2)\Big\} + 2\sum_{j=0}^{m-1}\sum_{k=j+1}^{m} d_j d_k\Big\{\sum_{i=0}^{j}\sum_{l=0}^{k} a^*_{ij}a^*_{lk}\theta^{i+l+1}\Gamma(i+l+1/\theta+1)\Big\}\Big]\frac{1}{c(d)\Gamma(1/\theta)}.$$

The proof of Lemma 2 is a straightforward application of the definition of an expected value and the definition of the Gamma function (see Web Appendix A).

Since $E(\tilde{U}_m)$ depends on the parameter vector $d$ and on the value of $m$, and because the frailties act in a multiplicative way on the baseline hazard in model (1), a standardization of the frailties to have mean equal to one, guarantees that model (1) is well defined. Indeed, otherwise the baseline hazard would change from one model to another. It readily follows that the density of the standardized frailty random variable $U_m = \tilde{U}_m/E(\tilde{U}_m)$, where $\tilde{U}_m$ has density $f_{\tilde{U}_m}$, equals

$$f_{U_m}(u) = f_U(uE_m)\{E_m/c(d)\}\{\sum_{j=0}^{m} d_j v_j(uE_m)\}^2. \tag{3}$$

## 3. The marginal log-likelihood

An important aspect of fitting data with the extended gamma frailty densities is that the marginal likelihood of the data has a closed form. With $H_0$ the cumulative baseline hazard,

define for cluster $s = 1, \ldots, S$,

$$A_s = \prod_{j=1}^{n_s} \{h_0(y_{sj}) \exp(x_{sj}^T \beta)\}^{\delta_{sj}}, \ B_s = \sum_{j=1}^{n_s} H_0(y_{sj}) \exp(x_{sj}^T \beta), \ D_s = \sum_{j=1}^{n_s} \delta_{sj}.$$

We denote by $\xi$ the parameter vector used to model the baseline hazard, and use $\zeta_m$ as notation for the vector $(\xi, \beta, \theta, d)$.

For cluster $s$, the likelihood of the data conditional on the frailties is $L_s(\xi, \beta | U_{m,s} = u_s) = u_s^{D_s} A_s \exp(-u_s B_s)$.

THEOREM 1: *The marginal log-likelihood of the data in the extended gamma frailty model with frailty density $f_{U_m}$ of* (3) *is with the above notation given by*

$$\begin{aligned}\ell_{m,\mathrm{marg}}(\zeta_m) &= \sum_{s=1}^{S} \log \int_0^\infty L_s(\xi, \beta | U_{m,s} = u_s) f_{U_m}(u_s) du_s \\ &= \sum_{s=1}^{S} \log\Big( \frac{A_s E_m^{1/\theta}}{c(d)\theta^{\frac{1}{\theta}}\Gamma(\frac{1}{\theta})} \Big[ \sum_{j=0}^{m} d_j^2 \Big\{ \sum_{i=0}^{j} \frac{b_{ij}^* E_m^{2i} \Gamma(2i + D_s + \frac{1}{\theta})}{(B_s + E_m/\theta)^{2i + D_s + \frac{1}{\theta}}} + \sum_{i=0}^{j-1} \frac{c_{ij}^* E_m^{2i+1} \Gamma(2i + D_s + \frac{1}{\theta} + 1)}{(B_s + E_m/\theta)^{2i + D_s + \frac{1}{\theta} + 1}} \Big\} \\ &\quad + 2 \sum_{j=0}^{m-1} \sum_{k=j+1}^{m} d_j d_k \sum_{i=0}^{j} \sum_{l=0}^{k} \frac{a_{ij}^* a_{lk}^* E_m^{i+l} \Gamma(i + l + D_s + \frac{1}{\theta})}{(B_s + E_m/\theta)^{i + l + D_s + \frac{1}{\theta}}} \Big] \Big).\end{aligned}$$

The proof of Theorem 1 is given in the appendix.

## 4. Order selection tests for gamma frailties

### 4.1 *Null and alternative hypotheses*

In model (1) we wish to test the null hypothesis that the frailty follows a gamma distribution with (unknown) shape parameter $1/\theta$ and scale parameter $\theta$, thus having mean one and variance $\theta$,

$$\mathcal{H}_0 : U \sim \Gamma(1/\theta, \theta), \text{ for some value } \theta > 0. \tag{4}$$

The alternative hypothesis $\mathcal{H}_a$ states that the frailty follows a distribution different from the gamma distribution. Using the extended gamma family, where models with different values of the truncation point $m = 0, 1, \ldots, M$ are fit to the data, the hypotheses are rephrased as

$\mathcal{H}_0$: $m = 0$ which is equivalent with: for all $j = 1, \ldots, M : d_j = 0$

$\mathcal{H}_a$: $m > 0$ which is equivalent with: there exists a $j \in \{1, \ldots, M\} : d_j \neq 0$.

Different values of $m = 1, 2, \ldots, M$ lead to different extensions of the gamma density. Models with a large value of the order $m$ contain the models with a smaller value of $m$ as special cases. In other words, a nested model sequence is constructed by letting $m$ grow.

### 4.2 *Likelihood ratio order selection tests*

For likelihood based models, Aerts et al. (1999) defined the order selection (OS) statistic by, rephrased to this setting,

$$T_{N,OS} = \max_{1 \leqslant m \leqslant M} 2 \, \frac{\ell_{m,\mathrm{marg}}(\widehat{\zeta}_m) - \ell_{0,\mathrm{marg}}(\widehat{\zeta}_0)}{m},$$

where $\zeta_m$ is the parameter vector of the model using density $f_{U_m}$, which is estimated under the alternative models by $\widehat{\zeta}_m$, while $\widehat{\zeta}_0$ is the estimator under the null model with a gamma frailty effect. The subscript $N$ denotes the finite sample aspects of the test statistic and is an indication of the size of the dataset. The length of $\zeta_m$ is denoted by $q_m$.

Expanding the extended gamma frailty model when $m$ increases with one unit implies adding one polynomial $v_{m+1}$ to the series, thus adding $d_{m+1}$ to the vector of coefficients. The denominator of $T_{N,OS}$ gives the difference in the number of parameters of model $m$ as compared to the null model, $q_m - q_0 = m$.

The omnibus, nonparametric, nature of the test becomes clear. The test is not a single likelihood ratio test, but a maximum of weighted likelihood ratio statistics. By taking a maximum, the statistic $T_{N,OS}$ combines several separate likelihood ratio statistics and avoids multiple testing issues. The weights take the complexity of the models into account, and down-weight large models. The use of a nonparametric series expansion avoids the specification of a parametric alternative model to the hypothesized null model.

The name 'order selection' becomes clear by rewriting the test that rejects when $T_{N,OS}$

is larger than a critical value $C_\alpha$ at significance level $\alpha$ in an equivalent form in terms of a modified version of Akaike's information criterion (AIC, Akaike, 1973) to select the order $m$. When modifying the traditional AIC by changing the penalty constant 2 to the value $C_\alpha$,

$$\mathrm{AIC}_{C_\alpha}(m) = 2\ell_{m,\mathrm{marg}}(\widehat{\zeta}_m) - C_\alpha q_m,$$

it is immediately clear that rejecting $\mathcal{H}_0$ when $T_{N,OS} > C_\alpha$ is equivalent to reject $\mathcal{H}_0$ when $\widehat{m} = \arg\max_{m=1,\ldots,M} \mathrm{AIC}_{C_\alpha}(m) > 0$.

### 4.3 *Asymptotic distribution and bootstrap resampling*

The asymptotic distribution of $T_{N,OS}$ coincides with that of the order selection statistic studied by Aerts et al. (1999) for parametric likelihood models. Since the series construction results in a sequence of nested models, the likelihood ratio statistics follow asymptotically a $\chi^2$ distribution with degrees of freedom equal to the difference in the number of parameters that is estimated in both models. Hence, the limiting form of $T_{N,OS}$ is, for $M \to \infty$,

$$T_{OS} = \max_{m \geqslant 1} \frac{V_m}{m}, \text{ where } V_m = \sum_{j=1}^{m} Z_j^2,$$

with $Z_1, Z_2, \ldots$ independent $N(0,1)$ distributed random variables. The value of $C_\alpha$ can easily be simulated from the asymptotic distribution. In particular, $C_{0.05} = 4.18$.

For finite samples our numerical study (see Figure B.1 in the Web Appendix) has revealed that the dependence of the orthogonal polynomials on the value of $\theta$ (see Lemma 1) is observable in the finite sample distribution of the test statistic, which, however, should disappear in the limit. For this reason we suggest a bootstrap resampling scheme to get more adequate results. For bootstrap results to be valid for testing, bootstrap data are generated under the null hypothesis. The parameter estimates obtained under the assumption of a one-parameter gamma frailty distribution are used to generate new event times and censoring times. More details are given in Section 5.

## 5. A small numerical performance study

To illustrate the performance of the proposed methodology, we use a small simulation study. The simulation settings are along the lines of the extensive simulation on the performance of the gamma frailty model in Section 5.2.3 of Duchateau and Janssen (2008).

**Event times**. The $T_{sj}$'s (in years) are generated using the shared frailty model (1). We take a Weibull baseline hazard with $\xi = (\lambda, \rho) = (0.22, 1)$ (exponential) and, for a binary covariate, $\beta = \log(1.3)$, i.e., a hazard rate of 1.3. The frailty densities we use are:

(i) $f_1$ is one-parameter gamma with $\theta = 0.3$;

(ii) $f_2$ is inverse Gaussian, i.e., $f_2(u) = (\alpha/2\pi)^{1/2} u^{-3/2} \exp((-\alpha/2u\mu^2)(u-\mu)^2)$ with $\mu = 2$ and $\alpha = 5$;

(iii) $f_3$ is a $30:70$ mixture of $f_1$ and $f_2$;

(iv) $f_4$ is a $70:30$ mixture of $f_1$ and $f_2$.

The frailty distributions in (ii)-(iv) serve as alternative frailty distributions.

**Censoring times** $C_{sj}$. Think about a trial where patients enter the study in a uniform way over an accrual period of five years and a follow up period of three years. The censoring time (time at risk) for a subject thus consists of the time at risk before the end of the accrual period plus the follow up time.

**Number of clusters**. $S = 150$ or $S = 300$.

**Number of observations per cluster**. $n_s \equiv n = 4$.

We further assume that we observe the minimum of the event time and the time at risk, $Y_{sj} = \min(T_{sj}, C_{sj})$. For simulating under the gamma frailty model, this resulted in about 33% censored observations. For the inverse Gaussian frailty distribution, the percentage of censored observations is approximately 15.

The concrete settings we consider for ($S$,$n$,frailty) (see Table 1) are (150,4,$f_1$), (300,4,$f_1$), (150,4,$f_2$), (300,4,$f_2$), (150,4,$f_3$), (150,4,$f_4$). To compute the bootstrap p-values in the sim-

ulation study we use the following algorithm (Massonnet et al., 2006; Davison and Hinkley, 1997).

**ALGORITHM** (given one of the six concrete settings)

**Step** 0. $r = 1$ ($r$ is the run index, $r = 1, \ldots, R$)

**Step** 1. Generate right censored data according to the above description of event/censoring time. Fit the loglikelihood in Theorem 1 with $m = 0, \ldots, M$ and obtain the actual value of the order selection test statistic: $T_{act}$ (we take $M = 3$). Let $\hat{\lambda}_0$, $\hat{\rho}_0$, $\hat{\beta}_0$ and $\hat{\theta}_0$ denote the parameter estimates for $m = 0$.

**Step** 2. Generate $B$ resamples ($b$ is the resample index, $b = 1, \ldots, B$).

**Step** 2.0. $b = 1$

**Step** 2.1. Sample $u_1^*, \ldots, u_S^*$ from a $\Gamma(1/\hat{\theta}_0, \hat{\theta}_0)$ distribution.

**Step** 2.2. For $j = 1, \ldots, n_s; s = 1, \ldots, S$, generate event times $T_{sj}^*$ from the estimated survival function

$$\hat{S}_{sj}(t) = \{\exp(-\hat{H}_0(t))\}^{u_s^* \exp(x_{sj}\hat{\beta}_0)}$$

with $\hat{H}_0(t)$ the estimated cumulative Weibull baseline hazard.

**Step** 2.3. If $\delta_{sj} = 0$ set $C_{sj}^* = Y_{sj}$; if $\delta_{sj} = 1$ generate $C_{sj}^*$ from a uniform distribution with the same accrual and follow up period as for the original data.

**Step** 2.4. Set $Y_{sj}^* = \min(T_{sj}^*, C_{sj}^*)$ with $\delta_{sj}^* = 1$ if $Y_{sj}^* = T_{sj}^*$; and $\delta_{sj}^* = 0$ otherwise.

**Step** 2.5. Obtain the bootstrap value of the order selection test statistic: $T_b^*$.

**Step** 2.6. $b \mapsto b + 1$ (until $b + 1 = B$).

**Step** 3. Obtain the bootstrap version of $P_{\mathcal{H}_0}(T > T_{act})$, i.e., $p_r^* = \#\{b : T_b^* > T_{act}\}/B$.

**Step** 4. $r \mapsto r + 1$ (until $r + 1 = R$).

Since the maximization of the likelihood is numerically difficult and time-consuming, especially when the value of $m$ is large, which combined with a bootstrap algorithm costs

even more time, we took $R = 100$ and $B = 150$. While for a single data example the accuracy can be taken higher, this was not feasible in a simulation study.

In Table 1 we list the percentage of times that the bootstrap version of the p-value is below the considered level of significance. In other words the percentage of times that the null hypothesis is rejected.

[Table 1 about here.]

Due to the Newton-type of algorithm that we used for the likelihood maximization, it happened a few times in the simulation study that the likelihood of a larger model was not larger than that of a smaller model, which has lead us to disregard these cases. Table 1 shows in its first two lines the results of the tests under the null hypothesis. Keeping in mind the size of the simulation study, this shows that the test is able to approximately reach the nominal significance level. The inverse Gaussian alternative model gives the highest simulated rejection probabilities amongst the considered cases. As expected, the simulated rejection probabilities of the test increases with increased sample size. The alternative models (iii) and (iv) are harder to detect, though the test is able to pick up the deviation from the null hypothesis situation, with, for the case of 150 clusters about equal performance.

## 6. Two illustrative examples

We use the proposed order selection statistic to test the null hypothesis that the frailty density is $\Gamma(1/\theta, \theta)$. First we fit the loglikelihood in Theorem 1 with $m = 0, \ldots, 5$ and obtain the actual value of the order selection test statistic. To obtain the bootstrapped p-value, we follow steps 2 and 3 of the algorithm described in Section 5, with this exception that we no longer assume that the censoring times come from a uniform distribution. Rather, we use a nonparametric Kaplan-Meier estimator of the censoring distribution. Particularly, we replace step 2.3 by

**Step** 2.3**(bis)**. If $\delta_{sj} = 0$ set $C^*_{sj} = Y_{sj}$; if $\delta_{sj} = 1$ generate $C^*_{sj}$ from the conditional censoring distribution given that $C_{sj} > Y_{sj}$, namely

$$\frac{\hat{G}(t) - \hat{G}(Y_{sj})}{1 - \hat{G}(Y_{sj})}$$

with $\hat{G}$ the Kaplan-Meier estimate of the censoring distribution (assume that $G$ is independent of the covariate).

### 6.1 *Analysis of the mastitis data*

Mastitis, an infection of the udder, is a disease in the dairy cow sector with an important economic impact. A cluster consists of the four udder quarters of a cow. In this example 100 cows are followed up for infections. We consider parity as the single (binary) covariate. The parity of a cow is the number of calvings (and therefore the number of lactation periods) that the cow has already experienced. Parity is often converted into a binary covariate, grouping all the cows with more than one calving in the group of multiparous cows (parity = 0) compared to the group of primiparous cows, cows with only one calving (parity =1). A few data lines are given in Table 2.

[Table 2 about here.]

We assume a Weibull baseline hazard, i.e., $\xi = (\lambda, \rho)$. Empirical evidence for the assumed Weibull baseline hazard is obtained by checking whether the population density, which is obtained by integrating out the frailty, provides a reasonable fit for the event times, see Figure 2.

The bootstrapped p-value (based on 300 resamples) is 0.69, we therefore do not reject the null hypothesis.

### 6.2 *Analysis of the insemination data*

Time from parturition to first insemination is one of the main factors determining the calving interval (the time between two calvings) which is optimally between 12 and 13 months. One

objective of artificial insemination programmes is to look for cow factors that might predict the time for parturition to first insemination, so that action can be taken based on these predictors. The concrete data set contains data from 181 dairy farms (= herds), the number of cows within a herd varies from 1 to 174, with an average of 58 cows. As in the previous example we consider parity as the single binary covariate. A few data lines are given in Table 3.

[Table 3 about here.]

Also for this example a Weibull baseline hazard is feasible, see Figure 2. For this example we obtain 0 as bootstrapped p-value (based on 157 resamples), i.e., there is no support for the null hypothesis that the frailty terms follow a $\Gamma(1/\theta, \theta)$ distribution. Due to the large cluster sizes in this data example, there were some more convergence problems of the optimization procedures. For this data set, the extended gamma frailty density with $m = 2$ was preferred by the Akaike information criterion. Figure 3 gives a graphical representation of the estimated null density and the estimated frailty density with $m = 2$.

[Figure 2 about here.]

[Figure 3 about here.]

## 7. Discussion and possible extensions

### 7.1 *Score-based tests*

A disadvantage of this likelihood-based test is that for each value of $m = 0, \ldots, M$, the model needs to be refit in order to find the maximum of the likelihood. An alternative to a likelihood ratio statistic is a score statistic. This has the advantage that the model only needs to be fit under the null hypothesis. For linear mixed models, the use of a score statistic for testing the distribution of random effects has been suggested by Thas (2009) in a discussion of Claeskens and Hart (2009). A limited simulation comparison has shown good

power behavior in comparison with the nested models based order selection test. For the shared frailty models, the score statistic requires the computation of the matrix of second partial derivatives with respect to all unknown parameters. If the fitting times are an issue, this might be worthwhile to derive.

### 7.2 *Singleton tests*

Another option to reduce computation times of a full likelihood-based test is to consider the so-called singleton expansions in which case we construct extensions of the gamma density that contain only a single term of the series expansion in the following way, $\check{f}_{U_j}(u) = f_U(u)v_j^2(u)$. Since the polynomials $v_j$ are orthonormal with respect to the gamma density, the integral of $\check{f}_{U_j}$ is one, and the square guarantees that the function is positive. Such singleton alternative models were considered by Aerts et al. (2004) in a regression setting. A difference with the above construction is that the different models are no longer nested, with the exception of the null model that is nested within each other model. Based on the set of singleton alternatives, we consider a test statistic that combines the likelihood values of the different singleton expansions of the gamma density.

$$T_{N,\text{singleton}} = \max_{m=1,\ldots,M} \{\log \check{L}_m(\hat{\xi}_m, \hat{\beta}_m, \hat{\theta}_m) - \log \check{L}_0(\hat{\xi}_0, \hat{\beta}_0, \hat{\theta}_0)\},$$

where $\check{L}_m$ $(m = 1, \ldots, M)$ denotes the likelihood of the model using the singleton alternative density for the frailty effect in the model. Also for this test a bootstrap procedure would be advised rather than working with the asymptotic distribution.

### 7.3 *Parametric frailty models: the choice of the baseline hazard*

Although we used a Weibull baseline hazard in the previous sections, the proposed methodology is valid for any parametric choice of baseline hazard, e.g., a piecewise constant baseline hazard.

7.4 *Semi-parametric frailty models: extending the penalized partial likelihood approach*

A challenging problem is to extend the results in this paper to frailty models with unspecified baseline (semi-parametric frailty models). We conjecture that for this problem the natural approach will be based on the penalized partial likelihood method (Therneau and Grambsch, 2000; Duchateau and Janssen, 2008).

7.5 *Testing for other frailty distributions*

The proposed orthogonal polynomials are specifically constructed for the gamma frailty distribution. Calculations using a similar type of arguments could be done for other distributions, but this is beyond the scope of the current paper. While the orthogonalization might still be explicit, there is no guarantee that for other distributions the marginal likelihood can be explicitly obtained.

## Supplementary Material

Web Appendix A, referenced in Section 2, is available under the Paper Information link at the *Biometrics* website `http:www.tibs.org/biometrics`. Web Appendix B contains an additional figure corresponding to the simulation output and is mentioned in Section 4.3. Web Appendix C contains the R code.

## Acknowledgements

The authors wish to thank A. Davison for helpful discussions and suggestions. This work was supported by the IAP Research Network P6/03 of the Belgian State (Belgian Research Policy). The mastitis data are retrieved from: http://www.vetstat.ugent.be/research/frailty/datasets/?dataset=0004mastitis; the time to first insemination data are obtained from: http://www.vetstat.ugent.be/education/survivalpartim2/0008insem.dat

## References

Aerts, M., Claeskens, G., and Hart, J. D. (1999). Testing the fit of a parametric function. *Journal of the American Statistical Association* **94,** 869–879.

Aerts, M., Claeskens, G., and Hart, J. D. (2000). Testing lack of fit in multiple regression. *Biometrika* **87,** 405–424.

Aerts, M., Claeskens, G., and Hart, J. D. (2004). Bayesian-motivated tests of function fit and their asymptotic frequentist properties. *The Annals of Statistics* **32,** 2580–2615.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In Petrov, B. and Csáki, F., editors, *Second International Symposium on Information Theory*, pages 267–281. Akadémiai Kiadó, Budapest.

Claeskens, G. and Hart, J. D. (2009). Goodness-of-fit tests in mixed models. *Test* **18,** 213–239. With discussion.

Davison, A. and Hinkley, D. (1997). *Bootstrap methods and their applications.* Cambridge University Press.

Duchateau, L. and Janssen, P. (2008). *The Frailty Model.* Springer, New York.

Duchateau, L., Opsomer, G., Dewulf, J., and Janssen, P. (2005). The nonlinear effect (determined by the penalised partial likelihood approach) of milk-protein concentration on time to first insemination in belgian dairy cows. *Preventive Veterinary Medicine* **68,** 81–90.

Eubank, R. L. and Hart, J. D. (1992). Testing goodness-of-fit in regression via order selection criteria. *The Annals of Statistics* **20,** 1412–1425.

Gallant, A. R. and Nychka, D. W. (1987). Semi-nonparametric maximum likelihood estimation. *Econometrica* **55,** 363–390.

Goethals, K., Ampe, B., Berkvens, D., Laevens, H., Janssen, P., and Duchateau, L. (2009). Modelling interval-censored, clustered cow udder quarter infection times through the

shared gamma frailty model. *JABES* **14,** 1–14.

Hart, J. D. (1997). *Nonparametric Smoothing and Lack-of-fit Tests.* Springer-Verlag, New York.

Massonnet, G., Burzykowski, T., and Janssen, P. (2006). Resampling plans for frailty models. *Communications in Statistics-Simulation and Computation* **35,** 497–514.

Thas, O. (2009). Comments on: Goodness-of-fit tests in mixed models. *Test* **18,** 260–264.

Therneau, T. and Grambsch, P. (2000). *Modeling Survival Data. Extending the Cox Model.* Springer, New York.

Zhang, D. and Davidian, M. (2001). Linear mixed models with flexible distributions of random effects for longitudinal data. *Biometrics* **57,** 795–802.

## Appendix: Proof of Theorem 1

With $D_s = \sum_{j=1}^{n_s} \delta_{sj}$ and $\prod_{j=1}^{n_s} u_s^{\delta_{sj}} = u_s^{D_s}$, the conditional likelihood for the s-th cluster is equal to

$$L_s\left(\xi, \beta | U_{m,s} = u_s\right) = u_s^{D_s} \prod_{j=1}^{n_s} \left[ \left\{ h_0\left(y_{sj}\right) \exp\left(x_{sj}^T \beta\right) \right\}^{\delta_{sj}} \exp\left\{ -H_0\left(y_{sj}\right) u_s \exp\left(x_{sj}^T \beta\right) \right\} \right].$$

Next we compute the marginal likelihood for the s-th cluster,

$$\begin{aligned} L_{m,marg,s}\left(\zeta_m\right) &= \int_0^\infty L_s\left(\xi, \beta | U_{m,s} = u_s\right) f_{U_m}(u_s) du_s \\ &= \prod_{j=1}^{n_s} \left\{ h_0(y_{sj}) \exp(x_{sj}^T \beta) \right\}^{\delta_{sj}} \int_0^\infty u_s^{D_s} \prod_{j=1}^{n_s} \left[ \exp\left\{ -H_0(y_{sj}) u_s \exp(x_{sj}^T \beta) \right\} \right] f_{U_m}(u_s) du_s. \end{aligned}$$

Denote $A_s = \prod_{j=1}^{n_s} \left\{ h_0(y_{sj}) \exp(x_{sj}^T \beta) \right\}^{\delta_{sj}}$ and $B_s = \sum_{j=1}^{n_s} \left\{ H_0(y_{sj}) \exp(x_{sj}^T \beta) \right\}$, then,

$$\begin{aligned} L_{m,marg,s}\left(\zeta_m\right) &= A_s \int_0^\infty u_s^{D_s} \exp(-u_s B_s) f_{U_m}(u_s) du_s \\ &= A_s \int_0^\infty u_s^{D_s} \exp(-u_s B_s) \frac{f_U(u_s E_m)}{c(d)} E_m \Big\{ \sum_{j=0}^{m} d_j v_j(u_s E_m) \Big\}^2 du_s \\ &= \frac{A_s E_m}{c(d)} \Big[ \int_0^\infty u_s^{D_s} \exp(-u_s B_s) f_U(u_s E_m) \Big\{ \sum_{j=0}^{m} d_j^2 v_j^2(u_s E_m) \end{aligned}$$

$$
\begin{aligned}
&+2\sum_{j=0}^{m-1}\sum_{k=j+1}^{m} d_j d_k v_j(u_s E_m) v_k(u_s E_m)\Big\} du_s\Big] \\
= \;& \frac{A_s E_m}{c(d)}\Big\{ \sum_{j=0}^{m} d_j^2 \int_0^\infty u_s^{D_s} \exp(-u_s B_s) f_U(u_s E_m) v_j^2(u_s E_m) du_s \\
&+2\sum_{j=0}^{m-1}\sum_{k=j+1}^{m} d_j d_k \int_0^\infty u_s^{D_s} \exp(-u_s B_s) f_U(u_s E_m) v_j(u_s E_m) v_k(u_s E_m) du_s\Big\} \\
= \;& \frac{A_s E_m}{c(d)}(G_1 + G_2).
\end{aligned}
$$

We explicitly compute the terms $G_1$ and $G_2$. Using the change of variable $t_s = u_s\left(B_s + \frac{E_m}{\theta}\right)$,

$$
\begin{aligned}
G_1 \;=\; & \sum_{j=0}^{m} d_j^2\Big\{ \int_0^\infty \exp(-u_s B_s) f_U(u_s E_m)\Big( \sum_{i=0}^{j} b_{ij}^* E_m^{2i} u_s^{2i+D_s} + \sum_{i=0}^{j-1} c_{ij}^* E_m^{2i+1} u_s^{2i+D_s+1}\Big) du_s\Big\} \\
= \;& \frac{E_m^{\frac{1}{\theta}-1}}{\theta^{\frac{1}{\theta}}\Gamma\left(\frac{1}{\theta}\right)} \sum_{j=0}^{m} d_j^2 \int_0^\infty \exp\{-u_s(B_s + E_m/\theta)\}\Big( \sum_{i=0}^{j} b_{ij}^* E_m^{2i} u_s^{2i+D_s+\frac{1}{\theta}-1} \\
&+ \sum_{i=0}^{j-1} c_{ij}^* E_m^{2i+1} u_s^{2i+D_s+\frac{1}{\theta}}\Big) du_s \\
= \;& \frac{E_m^{\frac{1}{\theta}-1}}{\theta^{\frac{1}{\theta}}\Gamma\left(\frac{1}{\theta}\right)} \sum_{j=0}^{m} d_j^2\Big\{ \sum_{i=0}^{j} \frac{b_{ij}^* E_m^{2i}}{(B_s + E_m/\theta)^{2i+D_s+\frac{1}{\theta}}} \int_0^\infty \exp(-t_s) t_s^{2i+D_s+\frac{1}{\theta}-1} dt_s \\
&+ \sum_{i=0}^{j-1} \frac{c_{ij}^* E_m^{2i+1}}{(B_s + E_m/\theta)^{2i+D_s+\frac{1}{\theta}+1}} \int_0^\infty \exp(-t_s) t_s^{2i+D_s+\frac{1}{\theta}} dt_s\Big\} \\
= \;& \frac{E_m^{\frac{1}{\theta}-1}}{\theta^{\frac{1}{\theta}}\Gamma\left(\frac{1}{\theta}\right)} \sum_{j=0}^{m} d_j^2\Big\{ \sum_{i=0}^{j} \frac{b_{ij}^* E_m^{2i}\Gamma\left(2i + D_s + 1/\theta\right)}{(B_s + E_m/\theta)^{2i+D_s+\frac{1}{\theta}}} + \sum_{i=0}^{j-1} \frac{c_{ij}^* E_m^{2i+1}\Gamma\left(2i + D_s + 1/\theta + 1\right)}{(B_s + E_m/\theta)^{2i+D_s+\frac{1}{\theta}+1}}\Big\}.
\end{aligned}
$$

For $G_2$ we use the same change of variables to obtain that

$$
\begin{aligned}
G_2 \;=\; & 2\sum_{j=0}^{m-1}\sum_{k=j+1}^{m} d_j d_k \int_0^\infty \exp(-u_s B_s) f_U(u_s E_m) \sum_{i=0}^{j}\sum_{l=0}^{k} a_{ij}^* a_{lk}^* E_m^{i+l} u_s^{i+l+D_s} du_s \\
= \;& \frac{2E_m^{\frac{1}{\theta}-1}}{\theta^{\frac{1}{\theta}}\Gamma\left(\frac{1}{\theta}\right)} \sum_{j=0}^{m-1}\sum_{k=j+1}^{m} d_j d_k \int_0^\infty \exp(-u_s(B_s + \frac{E_m}{\theta})) \sum_{i=0}^{j}\sum_{l=0}^{k} a_{ij}^* a_{lk}^* E_m^{i+l} u_s^{i+l+D_s+\frac{1}{\theta}-1} du_s \\
= \;& \frac{2E_m^{\frac{1}{\theta}-1}}{\theta^{\frac{1}{\theta}}\Gamma\left(\frac{1}{\theta}\right)} \sum_{j=0}^{m-1}\sum_{k=j+1}^{m} d_j d_k \sum_{i=0}^{j}\sum_{l=0}^{k} \frac{a_{ij}^* a_{lk}^* E_m^{i+l}}{\left(B_s + \frac{E_m}{\theta}\right)^{i+l+D_s+\frac{1}{\theta}}} \int_0^\infty \exp(-t_s) t_s^{i+l+D_s+\frac{1}{\theta}-1} dt_s \\
= \;& \frac{2E_m^{\frac{1}{\theta}-1}}{\theta^{\frac{1}{\theta}}\Gamma\left(\frac{1}{\theta}\right)} \sum_{j=0}^{m-1}\sum_{k=j+1}^{m} d_j d_k \sum_{i=0}^{j}\sum_{l=0}^{k} \frac{a_{ij}^* a_{lk}^* E_m^{i+l}\Gamma\left(i + l + D_s + 1/\theta\right)}{\left(B_s + \frac{E_m}{\theta}\right)^{i+l+D_s+\frac{1}{\theta}}}.
\end{aligned}
$$

Therefore,

$$\begin{aligned} L_{m,marg,s}\left(\zeta_m\right) &= \frac{A_s E_m^{\frac{1}{\theta}}}{c(d)\theta^{\frac{1}{\theta}}\Gamma\left(\frac{1}{\theta}\right)}\Big[\sum_{j=0}^{m} d_j^2\Big\{\sum_{i=0}^{j}\frac{b_{ij}^* E_m^{2i}\Gamma\left(2i+D_s+1/\theta\right)}{\left(B_s+\frac{E_m}{\theta}\right)^{2i+D_s+\frac{1}{\theta}}} \\ &\quad +\sum_{i=0}^{j-1}\frac{c_{ij}^* E_m^{2i+1}\Gamma\left(2i+D_s+1/\theta+1\right)}{\left(B_s+\frac{E_m}{\theta}\right)^{2i+D_s+\frac{1}{\theta}+1}}\Big\} \\ &\quad +2\sum_{j=0}^{m-1}\sum_{k=j+1}^{m} d_j d_k \sum_{i=0}^{j}\sum_{l=0}^{k}\frac{a_{ij}^* a_{lk}^* E_m^{i+l}\Gamma\left(i+l+D_s+1/\theta\right)}{\left(B_s+\frac{E_m}{\theta}\right)^{i+l+D_s+\frac{1}{\theta}}}\Big]. \end{aligned}$$

Taking the log and summing over the S clusters one gets the expression for the marginal loglikelihood $\ell_{m,marg}\left(\zeta_m\right)=\sum_{s=1}^{S}\log\left(L_{m,marg,s}\left(\zeta_m\right)\right)$ as stated in Theorem 1.

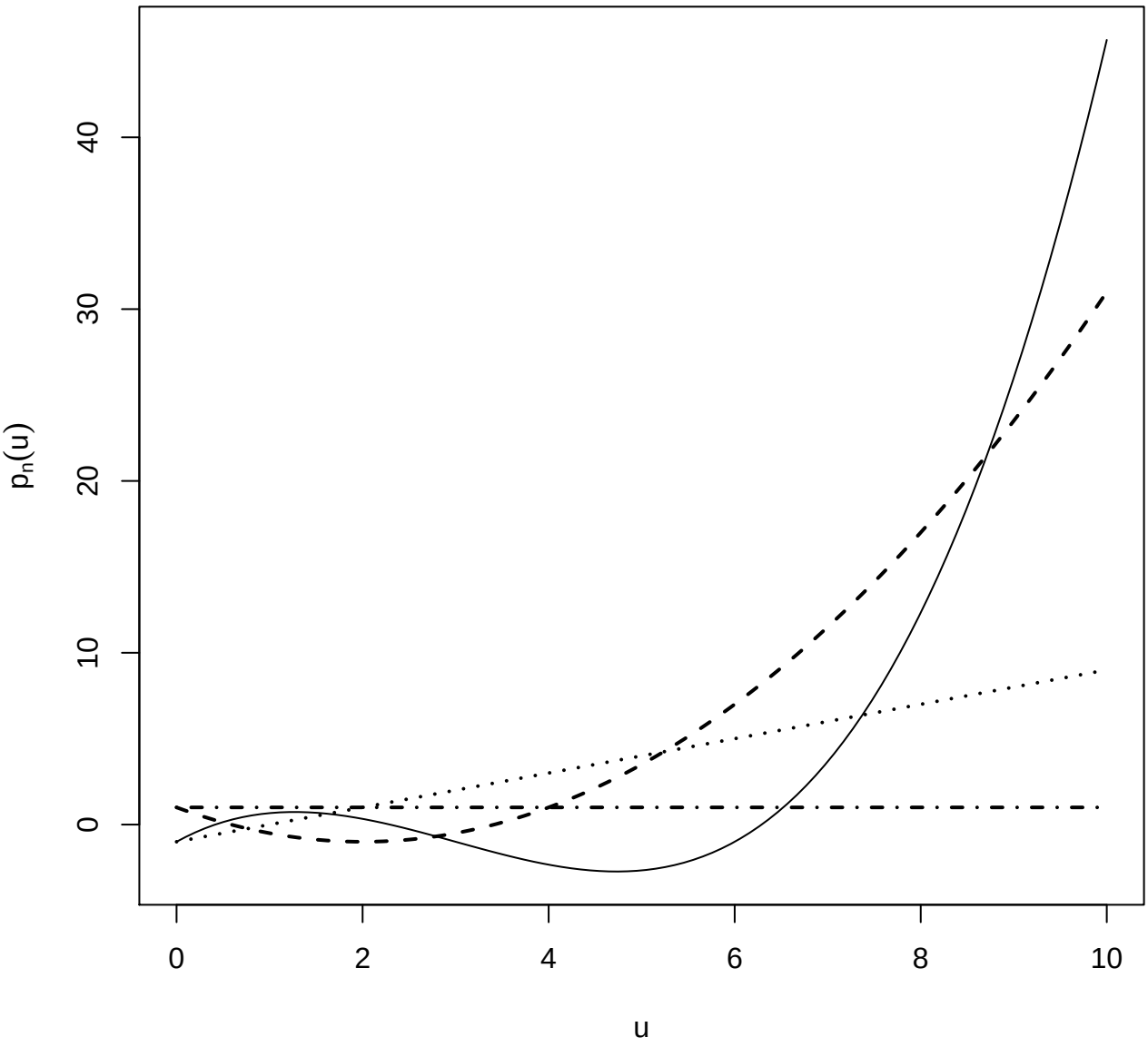


**Figure 1.** Polynomials orthogonal to the one parameter gamma density function with $\theta = 1$. Polynomial $p_0$ as a dot-dashed line, $p_1$ dotted, $p_2$ dashed and $p_3$ as a solid line.

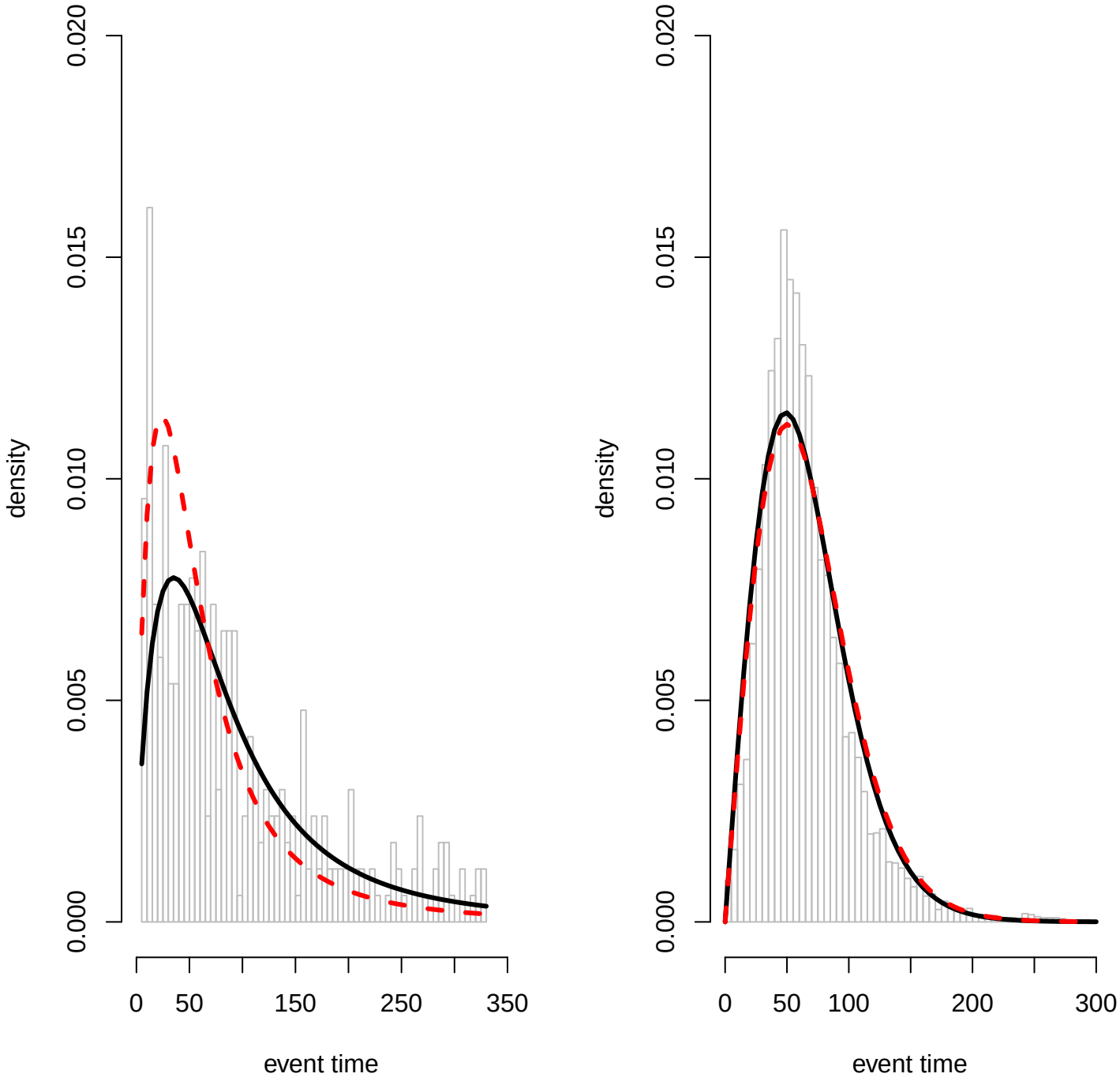


**Figure 2.** Estimated population density of the event times under the assumption of a Weibull baseline hazard (parity=0: solid line; parity=1: dashed line). Left panel: mastitis data, right panel: insemination data.

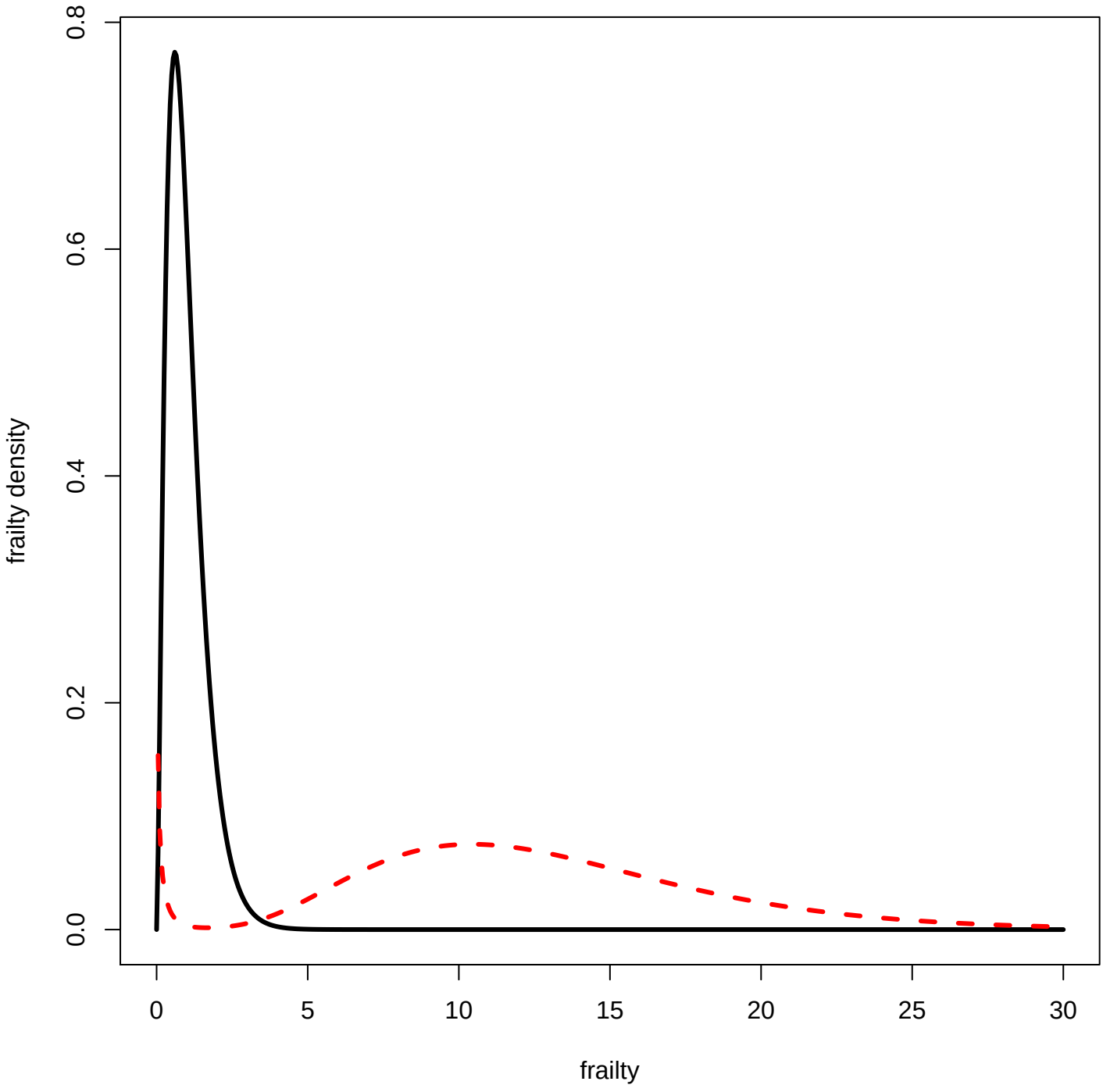


**Figure 3.** Graphical representation of the estimated null density, i.e., a one-parameter gamma density with $\hat{\theta} = 0.39$ (solid line) and the estimated selected density, i.e., an expanded gamma density with $\hat{\theta} = 2.64, \hat{d}_1 = 1.77$ and $\hat{d}_2 = 1.76$ (dashed line) for the insemination data.

**Table 1**

*Simulation results. Column* 1*: simulation setting; Column* 2*: number of simulated datasets for which the null hypothesis is rejected at level* $\alpha = 0.05$ *(percentage); Column* 3*: number of simulated datasets for which the null hypothesis is rejected at level* $\alpha = 0.10$ *(percentage)*

| ($S$,$n$,frailty) | $\alpha = 0.05$ | | $\alpha = 0.10$ | |
|---|---|---|---|---|
| (150,4,$f_1(.)$) | 7 out of 98 | (0.071) | 13 out of 98 | (0.133) |
| (300,4,$f_1(.)$) | 3 out of 95 | (0.032) | 11 out of 95 | (0.116) |
| (150,4,$f_2(.)$) | 10 out of 98 | (0.102) | 16 out of 98 | (0.163) |
| (300,4,$f_2(.)$) | 26 out of 98 | (0.265) | 37 out of 98 | (0.378) |
| (150,4,$f_3(.)$) | 10 out of 97 | (0.103) | 15 out of 97 | (0.155) |
| (150,4,$f_4(.)$) | 9 out of 97 | (0.093) | 15 out of 97 | (0.155) |

**Table 2**

*Mastitis data. Column* 1*: the cow identification number; Column* 2*: the minimum of the infection time (in days) and the censoring time for each of the four udder quarters (LF = left-front, LR = left-rear, RF = right-front and RR = rigth-rear); Column* 3*: the corresponding censoring indicators; Column* 4*: the parity (parity is* 1 *resp.* 0 *for primiparous resp. multiparous cows)*

| cow id | $\min(T_{sj}, C_{sj})$ (LF,LR,RF,RR) | censoring indicators | parity |
|---|---|---|---|
| 1 | (296,305,301,332) | (1,0,0,0) | 1 |
| 2 | (172,82,230,67) | (1,1,1,1) | 1 |
| ... | ... | ... | ... |
| 99 | (156,335,285,178) | (1,0,1,1) | 0 |
| 100 | (104,63,202,17) | (1,1,1,1) | 0 |

**Table 3**

*Time to first insemination. Column* 1*: the herd identification number (herd* 1 *has* 51 *cows, . . . , herd* 181 *has* 81 *cows); Column* 2*: the minimum of the time from parturition to first insemination (in days) and the censoring time; Column* 3*: the censoring indicator; Column* 4*: the parity (parity is* 1 *resp.* 0 *for primiparous resp. multiparous cows)*

| herd id | $\min(T_{sj}, C_{sj})$ | censoring indicator | parity |
|---|---|---|---|
| 1 | 68.5 | 1 | 0 |
| ... | ... | ... | ... |
| 1 | 70.5 | 1 | 1 |
| ... | ... | ... | ... |
| 181 | 48.5 | 1 | 0 |
| ... | ... | ... | ... |
| 181 | 155.5 | 1 | 1 |